Running head: SUBGROUP DIFFERENCES IN GBG IMPACT

For Whom does it Work?
Subgroup Differences in the Effects of a School-Based Universal Prevention Program

Jantine L. Spilt
Hans M. Koot
Pol A. C. van Lier

Department of Developmental Psychology
VU University Amsterdam, The Netherlands

Acknowledgements

The study was financially supported by ZonMW Grant 26200002 and NWO Grant 056-35-012.

Abstract

This study examined subgroup differences in the effectiveness of a universal classroom-based preventive intervention. The Good Behavior Game (GBG) was delivered in Grade 1 and 2 in a randomized controlled trial including 759 students. Changes in externalizing and internalizing problems were modeled from Kindergarten through Grade 2. Unlike previous research, a person-centered approach was employed to examine critical combinations of child-, peer-, family, and demographic characteristics at baseline as moderators of intervention impact. Six subgroups were identified that differed both in baseline risk profiles and intervention responsiveness. The GBG prevented the development of externalizing and internalizing behavior among low-risk children, children with emotional problems, and victimized children. No positive intervention effects were found for children from dysfunctional families and children with combinations of behavioral and social risk. The study presented a novel approach to study subgroup differences in universal preventive interventions and provides first evidence that universal school-based programs may not be effective for children with more severe risks and risks at multiple levels.

Keywords: Universal preventive intervention, randomized controlled trial, internalizing and externalizing behavior, person-centered approach

For Whom does it Work?

Subgroup Differences in the Effects of a School-Based Universal Prevention Program

Universal interventions are typically designed to prevent the development of psychopathology among an entire population. Many of such universal programs are delivered in schools targeting a wide range of maladaptive behaviors among students. However, research that has examined whether such programs are effective for all children or for specific subgroups of children only is scarce (Farrell, Henry, & Bettencourt, 2011). Interventions can work differently for children with different characteristics and developmental risks. It is therefore of primary importance to determine *for whom* universal interventions are effective (Flay et al., 2005). The overarching goal of the present study was to inform the role of universal school-based interventions in the prevention of mental health problems among children who differ in their exposure to various risk factors. More specifically, this study aimed to identify critical *combinations* of child, peer, family, and demographic characteristics that can explain subgroup differences in the effectiveness of the Good Behavior Game (GBG: Dolan, Werthamer, & Kellam, 1989); a school-based universal preventive intervention that targets behavioral and associated problems.

Who Profits from Universal Preventive School-Based Interventions?

While there are multiple studies in the clinical science literature that have examined moderators of intervention response (e.g., Beauchaine, Webster-Stratton, & Reid, 2005; Hinshaw, 2007), the issue has received far less attention in prevention science. This is surprising given that, unlike selective and indicated interventions, universal preventive programs target an entire population regardless of individual risk factors. The heterogeneous nature of the targeted population increases the likelihood of individual differences in intervention effects due to

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

variations in pre-existing individual and environmental characteristics. Though important efforts have been made in prevention science to establish main intervention effects and to unravel proximal processes behind these effects, the question *for whom* it works has remained largely unanswered. It is thus not only possible that intervention effects in fact are accounted for by only a limited subgroup of children, but also that beneficial effects for a particular subgroup of participants masks the *in*effectiveness or even adverse effects for others. For instance, research from the Multisite Violence Prevention Project (2008) demonstrated beneficial effects for high-risk students but an unwanted negative impact on low-risk students. Such findings underscore the necessity to move beyond analyses of main effects toward the identification of children who are most or least likely to respond positively to preventive interventions. This knowledge is not only important from a theoretical perspective but is also crucial for mental health school services to identify at-risk children in need of additional, more extensive support.

The extant literature that has examined moderator variables in prevention science is decidedly mixed. Most studies have examined gender as moderator of intervention effects. For instance, Bierman et al. (2010) found that boys benefited more from the Fast Track curriculum than girls when peer-rated outcomes were examined. This moderation effect, however, was not found when teacher reports were analyzed. In an earlier study, Bierman (2002) also found no evidence of gender moderated intervention effects on parent- and teacher-rated aggression. Farrell et al. (2011) conducted a comprehensive review of studies of subgroup effects in universal school-based violence prevention programs. They concluded that differences between boys and girls in intervention effects are inconsistent across studies. Moreover, differences between boys and girls in externalizing outcomes are most likely a result of boys exhibiting higher levels of disruptive behaviors than girls, as previous research on the GBG suggests (e.g., Van Lier, Vuijk,

& Crijnen, 2005; Witvliet, Van Lier, Cuijpers, & Koot, 2009). Also with respect to other variables at the individual child level (e.g., demographic characteristics, initial problem behaviors) and variables at the family level (only three studies), Farrell and colleagues concluded that there was no consistent pattern of moderation across studies. A possible explanation for the mixed results of this review is that attributes of the child and the social context are typically studied as isolated variables based on a variable-centered approach. However, risk factors rarely operate in isolation. Ecological and transactional models of development emphasize the co-occurrence and joint influences of multiple risk factors within and across domains and ecological levels of development (Bronfenbrenner, 1977; Ford & Lerner, 1992). We therefore aimed to select a broad range of moderator variables to identify a priori unknown subgroups of children with different risk profiles in order to be able to identify children who do profit from universal school-based interventions versus children who are either partial or nonresponders to such efforts.

The Good Behavior Game and the Study of Subgroup Effects

The GBG is a contingency program that offers a group reward system to promote appropriate student behavior and to prevent the development of behavior problems and associated problems. Through explicitly formulated classroom rules, an emphasis on positive behavior rewards instead of discipline, and the facilitation of positive peer interactions through a team-based approach, a predictable and emotionally-safe classroom climate is created. Rigorous RCT research demonstrates that the GBG has an significant and long-term impact on a wide range of mental health problems including externalizing problems, substance abuse, and internalizing problems, and that the GBG appears especially effective among aggressive children (e.g., Ialongo, Poduska, Werthamer, & Kellam, 2001; Kellam et al., 2008; Petras, Kellam, Brown, Muthén, Ialongo, & Poduska, 2008, Van Lier et al., 2005). Research investigating the underlying

mechanisms suggests that the GBG promotes peer acceptance and affiliations between typical and antisocial children, which accounts for reductions in antisocial behavior (Van Lier et al., 2005; Witvliet et al., 2009). However, no research has yet examined the effectiveness of the GBG in relation to baseline risk profiles reflecting different constellations of risks across developmental domains.

To facilitate the study of subgroup differences in responsiveness to universal prevention programs like the GBG, we suggest the following approach. With regard to the selection of risk variables it is important to focus on a set of factors that are believed to contribute to the onset, continuity or change in the developmental problems targeted by the GBG (Farrell et al., 2011). Second, the selected risk variables should represent multiple domains of development (child-, peer-, family-, and demographic influences) because it is the joint influence of risk factors across domains that likely explains differences in intervention responsiveness (cf. Loeber & Hay, 1997). Third, from a clinical perspective, the goal of studying subpopulations is the early identification of at-risk children who are not responsive to universal programs and are in need of additional support. Therefore, we suggest studying risk factors that can be easily assessed at school entry, for instance through a relatively short screening device. Finally, in the statistical approach of studying subpopulations with differential intervention effectiveness, we propose the employment of a person-centered approach (cf. Lanza & Rhoades, 2011). Whereas variable-centered approaches generally examine the unique effects of separate risk factors or the additive effects of multiple risks, person-centered approaches such as Latent Class Analyses (Muthén & Muthén, 2000) allow for an examination of not only the severity or strength of risk and protective factors but also the constellation of risk factors and the joint influences of these constellation of risk factors within and across various domains of development.

Selection of Risk Factors

At the *child level*, severity or level of dysfunction and possible comorbidity needs to be considered. First, the effectiveness of universal preventive intervention likely depends on the initial level of behavioral and emotional problems (Farrell et al., 2011). In case of severe and persistent dysfunction, a universal classroom program may not be able to foster the specific needs of high-risk children (cf. Van Lier et al., 2004). However, most research suggests that the impact of the GBG is highest among children with high initial levels of aggression (Kellam et al., 2008; Petras et al., 2008; Van Lier et al., 2005). In addition to severity, the co-occurrence of different behavior problems should be considered. For instance, when aggression co-occurs with attention problems and/or impulsivity or with internalizing disturbances, this may indicate more severe and persistent psychopathology (e.g., Loeber & Hay, 1997), suggesting limited effects of a unimodal classroom program.

At the *social level*, there is ample evidence that social problems with peers are a risk factor for the early development and continuity of both internalizing and externalizing problems (Ladd, 2006; Parker, Rubin, Erath, Wojslawowicz, & Buskirk, 2006). The GBG, however, may counteract these social risks by promoting affiliations between typical and at-risk children (Van Lier et al., 2005; Witvliet et al., 2009). In addition, the GBG could also help children with emotional disturbances who face increased risk of social exclusion or rejection by their peers (cf. Gazelle & Ladd, 2003).

The *family context* also plays an important role in the development of psychopathology. The nuclear family is the primary socializing context for developing children, which makes it unlikely for a unimodal classroom program to counteract the direct negative influences of poor parenting and parental stress on children's mental health (e.g., Ashford, Smit, Van Lier, Cuijpers,

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

& Koot, 2008; Patterson, DeBaryshe, & Ramsey, 1989; Shaw, Bell, & Gilliom, 2000). In addition, parenting practices have been found to moderate peer risk factors (e.g., Farrell et al., 2011), which underscores the importance of modeling joint influences of risks in multiple domains. Finally, *sociodemographic risks* (e.g., ethnic minority and low socioeconomic status) are believed to exert a broad influence on children's development via family processes and parenting quality (e.g., Dodge & Pettit, 1994) and to exacerbate family risk (Shaw et al., 2000).

Current Study

The aim of the current study was to identify a priori unknown subgroups of children with particular *combinations* of baseline characteristics that respond qualitatively differently to the Good Behavior Game (GBG). To this end, we studied intervention effects of the GBG on growth trajectories of behavioral and emotional problems from Kindergarten to Grade 2. The GBG was delivered in Grade 1 and 2. This study contributes to previous RCT research on the effectiveness of the GBG using the same sample and outcome measures (Witvliet et al., 2009). These previous studies reported declines in behavior problems and stability in emotional problems from Kindergarten to Grade 2 for intervention children, whereas significant growth in both behavioral and emotional problems was found among control children. This study extends previous research by identifying unobserved subgroup differences in intervention response rather than studying main effects. We expected that a preventive classroom contingency program like the GBG would demonstrate limited effects on the developmental trajectories of children with more severe dysfunction or comorbid psychopathology, of children exposed to risks at multiple domains, and of children with risks present at levels not targeted by the GBG, that is the family and demographic level.

Method

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

Sample

The sample consisted of 759 children from 47 classes in 30 regular elementary schools. Children were on average 6.0 years old (*SD* = .47) at the start of the study. Thirty-eight percent of the children were from low socioeconomic status (SES) families. The sample was ethnically-diverse: 57% Dutch/Caucasian, 11% Moroccan, 10% Turkish, 6% Surinam, 5% Netherlands Antillean, and 12% other. In the Netherlands, non-Western immigrant children are at increased risk of school and mental health problems (Stevens & Vollebergh, 2008). Informed consent was obtained from parents.

Classes were randomly assigned to the experimental (66%) or control condition. There were no differences between conditions with respect to children's sex, study dropout, and school location (i.e., section of the country and rural vs. urban) but children in the control condition were more often from low SES and ethnic minority families (for more details, see Witvliet et al., 2009). In the present study, only children with at least two of the four outcome assessments were included (*N* = 741; 51% boys). Inclusion/exclusion was not related to intervention condition (χ^2 [1, *N* = 759] = .11, *p* =.82).

Preventive Intervention

The GBG is an empirically-based universal classroom preventive interventions that aims to reduce disruptive behavior and promote social competencies (Dolan et al., 1989). It offers a reinforcement-based group management strategy. Short and long-term beneficial effects have been reported in multiple randomized controlled trials (RCTs) across countries, grade levels, and for different target behaviors including behavioral and emotional problems (e.g., Barrish, Saunders & Wolf, 1969; Ialongo et al., 1999/2001; Kellam, Rebok, Ialongo, & Mayer, 1994; Kellam et al., 2008; Van Lier et al., 2004/2005). To play the GBG, teachers assign children to

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

teams of 4-5 members with equal numbers of disruptive and non-disruptive children. Each team receives a number of cards that can be taken by the teacher when the team or a team member violates a predefined rule. Teams are encouraged to support each other in appropriate behavior that is rewarded by compliments. In addition, teams can win rewards when at least one card is left at the end of a 15-60 minute period. Teachers attended three training sessions and received ten annual classroom supervisions from licensed GBG supervisors (for details, see Witvliet et al., 2009).

Measures

The intervention was delivered in Grade 1 and 2. The outcome variables externalizing and internalizing behavior were assessed annually at four measurement occasions: spring of Kindergarten (0), spring of Grade 1 (1), fall Grade 2 (2) and early summer at the end of Grade 2 (3). Child and social risk variables were collected in Kindergarten. Data on parenting and demographic risk were collected in the spring of Grade 1.

Outcome Variables

Externalizing and internalizing behavior. The Problem Behavior at School Interview (PBSI; Erasmus Medical Center, 2000), administered by trained student research assistents, was used to assess externalizing and internaliznig behavior. In the structured interview, teachers were asked to rate pupils' behaviors on a 5-point Likert scale ranging from 0 (never applicable) to 4 (often applicable). Mean scores of the Conduct Problems scale ("this child starts fights"; 12 items; α = .90-.92) and the Oppositional Defiant Problems scale ("this child argues frequently"; 7 items; α = .89-.96) were averaged to create the Externalizing Behavior scale (cross-time range of *rs* between the ODD and CD = .83-.84). Internalizing Behavior was the mean score of the scales Anxiety ("Anxious", "Easily worried"; 5 items; α = .81-.83) and Depression ("Unhappy or

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

depressed"; 7 items; α = .78-.85). The PBSI has demonstrated adequate validity in multiple studies (e.g., Witvliet et al., 2009).

Risk Variables

Child Behavior Checklist – Teacher Report Form (TRF). Kindergarten teachers completed the subscales Externalizing problems (e.g., "Gets into many fights"; 32 items; α=.67), Attention problems (e.g., "Can't sit still, restless, or hyperactive"; 26 items; α=.93), Social problems ("Doesn't get along with other pupils"; 11 items; α=.73), and Internalizing problems (e.g., "Cries a lot"; 22 items; α=.70) of the TRF (Achenbach, 1991). Behaviors were rated on a 3-point scale (0= not true; 2= very true/ often true). The Dutch translated TRF demonstrates adequate psychometric properties (Verhulst, Van der Ende, & Koot, 1997).

Victimization. Relational victimization ("Is being excluded when a classmate is angry at him/her"; 3 items; α=.90) and Physical victimization ("Is being hit and kicked by classmates"; 3 items; α=.83) were assessed with the PBSI (see above) and averaged into a single score (*r*=.49).

Maternal involvement. The Alabama Parenting Questionnaire (APQ; Shelton, Frick, & Wootton, 1996) is a widely-used measure of parenting quality. Mothers completed the 10-item subscale Involvement (e.g., "You play games or do other fun things with your child; α=.81) on a 5-point Likert scale (1=never, 5= always). Observational research has supported the validity of the subscale (Hawes & Dadds, 2006), for which, in addition, associations with both internalizing and externalizing behaviors have been reported (e.g., McKee et al., 2008).

Parenting stress. The Parent Domain of the Nijmegen Parenting Stress Index was employed to assess parenting distress (Abidin, 1983; De Brock, Vermulst, Gerris, & Abidin, 1992). Mothers rated 11 items (e.g., "Being a parent to this child is more difficult than I thought"; α =.76) on a 6-point Likert scale ranging from 0 (completely disagree) to 5 (completely agree).

Maternal depression. The subscale Depressed Mood of the K10 scale was used, which is a short screening measure designed to monitor population prevalence of psychological distress (Kessler, et al., 2002). Mothers rated 3 items ("How often do you feel somber or depressed?"; α=.82) on a 5-point scale (0=never; 4=always). Previous research reported significant associations of the measure with externalizing and internalizing problems (e.g., Kahn, Brandt, & Whitaker, 2004).

Demographic risks. Ethnicity was dummy coded into Minority status (0=Dutch; 1=Non Dutch). A dummy variable of Socioeconomic status was used based on maternal and paternal educational level and vocation (0=average to high; 1=low).

Statistical Approach

The model was analyzed using the Mplus program (Muthén & Muthén, 1998-2007). Mplus allows for the possibility to include Latent Class Analyses (LCA) into a growth mixture modeling (GMM) framework that includes risk profiles, intervention condition, and the intervention outcomes. The objective is to identify the smallest number of latent classes that adequately explains the associations among the variables. Latent classes or unobserved groups of individuals are identified with similar response patterns (i.e., baseline risk profiles) based on a number of observed dependent variables (i.e., latent class indicators) *and* similar growth trajectories. The estimated parameters in the model were: a) latent class membership, and for each class: b) class-specific baseline profiles, that is means of each latent class indicator, c) means of growth parameters (intercept and slope of Externalizing and Internalizing behavior) for each class, and d) an estimate of the effect of GBG on the slopes of Externalizing and Internalizing behavior. Initial data explorations showed no evidence of quadratic growth, therefore linear slopes were modeled. Intercept and slope estimates of the same outcome were

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

allowed to correlate; these correlations were set equal between classes. In addition, correlations between the intercepts and between the slopes of the Externalizing and Internalizing patterns were allowed. The variances of latent class indicators and growth parameters were also set equal between classes.

There are no strict criteria to assess the number of classes. Widely accepted guidelines are: a) smallest Bayesian information criteria (BIC) value, b) adequate classification of class membership as indexed by entropy scores close to 1, and c) reasonable class counts (Jung & Wickrama, 2008). In addition, class solution should be evaluated in terms of theoretical relevance, usefulness, and parsimoniousness (Muthén & Muthén, 2000). The bootstrapped likelihood ratio test (BLRT) has been recommended to avoid overextraction of classes (Nylund, Asparouhov, & Muthen, 2007). The BLRT compares the optimal *k* class solution with the *k-1* class solution.

Mplus offers the advantage to accommodate incomplete data using full information maximum likelihood (FIML) estimation under the assumption of missing at random (MAR) in order to retain the full sample. The covariance coverage matrix provides information on the extent of missing data. The covariance coverage ranged between .426 and .996 across all data points, which is well above the minimum threshold of .100 of Mplus. The nested structure of the data was taking into account using the cluster option of Mplus with TYPE=Complex routine and the MLR estimator.

Results

Table 1 reports the descriptives of and correlations between the continuous study variables. All correlations were in the expected directions.

Model Fitting

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

Models up to seven classes were explored to assess the optimal number of classes. The BIC values of the four-, five-, six-, and seven-class models were 12,849, 12,630, 12,485, and 12,442, respectively. The Entropy values of these models were .87, 88, 85, and 85, respectively. The drop in BIC value between the six- and seven-class models was minor. The BLRT test favored the six-class model above the five-class model ($p<.05$). For these reasons, the six-class model was selected as the final model. Latent class probabilities of the selected model ranged from .87-.96, which indicates that children were classified to one of the six classes with satisfactory precision.

Baseline risk profiles are represented in Figure 1. Table 2 reports standardized mean differences in observed scores between the Control and GBG condition after two years of intervention as an approximation of effect sizes. Table 3 presents the class sizes, class-specific growth parameters, and estimated intervention effects for each class on Externalizing and Internalizing behavior indicated by significant effects of slope on GBG. In comparison to the *Low risk group,* heightened intercept levels of both Externalizing and Internalizing behavior were observed for all other subgroups, indicating that these subgroups represented children at-risk of mental health problems.

Risk Profiles

Fifty-three percent of the children (42% boys) were classified into a low-risk profile *(Low risk)*. These children scored low on each of the baseline variables (Figure 1). Within this class, intervention condition was a significant predictor of change in both Internalizing and Externalizing behavior, suggesting positive intervention effects (Table 3). Fourteen percent of the children were victimized by their peers but did not exhibit heightened levels of other risk factors (*Victimization risk;* 56% boys). For this class, the intervention had a significant positive effect on

Internalizing behavior only. Eight percent of the children experienced emotional disturbances *(Internalizing risk;* 53% boys*)*. These children had the highest scores on Internalizing problems and modest scores on Social problems. Substantial positive intervention effects on both Internalizing and Externalizing behavior were found for this class. A fourth class consisted of children (14%; 60% boys) who experienced familial difficulties including high Parenting stress, maternal Depression, and low Involvement. Additionally, these children were often from low SES and ethnic minority families *(Family and demographic risk)*. Within this group, no significant differences in slopes of Externalizing and Internalizing behavior between intervention children and control children were found. Finally, two classes were found with behavioral and social problems, one displaying moderately-high levels and one showing high levels of risks. Moderately-high levels of behavioral and social problems were observed for 7% of the children *(Moderately-high sociobehavioral risk*, 71% boys*)*. These children scored medium high on Externalizing problems, Attention problems, Social problems, and Victimization (1 to 2 *SD* above the mean). No significant intervention effects were found for this class. A small group of 3% of the children (86% boys) evidenced the highest levels of social and behavioral risks *(Severe sociobehavioral risk)*. These children evidenced severe levels of Externalizing problems, Attention problems, and Social problems (2.5 to 4 *SD* above the mean). Additional risks included the highest level of Victimization and heightened levels of Internalizing problems (1 to 1.5 *SD* above the mean). Also, these children were somewhat more likely to come from low SES families. No significant intervention effects were found within this class.

Differences in Intervention Effects between Risk Profiles

A series of Wald tests of parameter constraints were conducted to test differences in intervention effects (slope regressed on intervention status) between latent classes (Table 3). In

comparison to the *Low risk group*, significantly stronger intervention effects on the development of Externalizing behavior were found in the *Internalizing risk group*. No significant differences were found between the *Low risk group* and the other risk groups, which could be due to the large inequality in the subsample sizes, which compromises statistical power to detect group differences. The intervention effects in the *Internalizing risk group* differed significantly from the effects found in the *Family and demographic risk group* and the *Moderately-high sociobehavioral risk group* but not from the effects found in the *Victimization risk group* and the *Severe sociobehavioral risk group*.

Significant class differences in intervention effects on the development of Internalizing behavior were found as well. In comparison to the *Low risk group*, significantly stronger intervention effects were established in the *Internalizing risk group*. The difference between the *Low risk group* and the *Victimization risk group* was marginally significant ($p = 0.07$). The other risk groups did not differ significantly from the *Low risk group.* The *Internalizing risk group* differed significantly from all other risk groups. The *Victimization risk group* differed significantly from the *Family and demographic risk group* and marginally significantly from the *Moderately-high sociobehavioral risk group* ($p = 0.08$).

Discussion

This study found subgroup differences in the effectiveness of the Good Behavior Game based on differences in baseline risk profiles at school entry. The results indicated that while this universal preventive classroom-based program is effective for children with internalizing problems, for victimized children, and for low-risk children, it may not be effective for children

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

with moderately-high to very high sociobehavioral risks and for children exposed to family and demographic risks.

Though universal school-based prevention programs typically aim at heterogeneous populations, this study is among the first in prevention science that considers the likelihood that intervention effects vary for subgroups characterized by combinations of pre-existing individual and environmental characteristics. It aligns with the recently renewed interest in the search for subgroup effects in preventive intervention (Bloom & Michalopoulos, 2011; Farrell et al., 2011; Greca, Silverman, & Lochman, 2009; Lochman, 2004, 2006). We examined critical combinations of baseline child, peer, family, and demographic characteristics as moderators of universal preventive intervention effects in a randomized controlled trial. The results suggested considerably variability in both baseline risk profiles and intervention effects.

About half of the sample could be classified into a low-risk profile at baseline. Control children with a low-risk profile showed growth in behavioral and emotional problems over time, but for low-risk children who received the GBG these increases were significantly attenuated. This suggests that the GBG is effective in preventing normative growth in behavioral and emotional problems. The other half of the sample was classified into five different risk profiles that represented various combinations and levels of risks. The children in these risk profiles evidenced heightened levels of behavioral and emotional problems at baseline. Two of the five risk groups were positively impacted by the GBG, whereas the children in the other three risk groups could be qualified as intervention non-responders.

The two risk groups that were positively impacted by the GBG were children with internalizing problems and children who were victimized in kindergarten. In both groups, intervention children evidenced substantial declines in both externalizing and internalizing

problems. The largest intervention effects were established for children with internalizing problems. Social-anxious and/or withdrawn-depressed children may have particularly profited from a more predictable and emotionally-safe classroom climate, resulting from an emphasis on collaboration and positive behavior rewards. In addition, considering the heightened social problems of emotionally-disturbed children, supportive interactions with typical children could have been enhanced by the program. Research has shown that social-anxious children who are excluded by peers, but not non-excluded social-anxious children, are at risk of developing clinical depression (Ladd & Gazelle, 2003). Thus, an increase in peer acceptance could be a likely explanation for the positive effects of the GBG found among victimized children (cf. Van Lier et al., 2005; Witvliet et al., 2009).

Two risk groups were found representing children with primarily externalizing problems in combination with attention problems, internalizing problems, and social problems. These risk profiles indicated a combination of serious behavioral, cognitive, and social risks, but differences in severity of risks (i.e., moderately high versus very high). Within these groups, no significant effects between the GBG children and the control children were found. Although tests of between-class differences did not indicate differences between these two groups and the positive intervention effects found among low-risk children, suggesting perhaps positive intervention effects, this probably should be attributed to reduced statistical power to detect between-group differences due to the large inequality in subsample sizes. A behavioral-environmental contingency program as the GBG is unlikely to target the multiple etiological mechanisms that underlie diverse and persistent psychopathological problems.

The results of this study seem not in line with those from previous studies that have reported the most beneficial effects of the GBG among high aggressive-disruptive children

(Kellam et al., 2008; Petras et al., 2008; Van Lier et al., 2005), though comparisons with other studies are difficult because of many differences in study design, sample, and analyses. One of the main differences between earlier research and the present study is that earlier studies have generally focused on a single baseline factor, that is initial differences in antisocial-disruptive behavior. This study, however, did not find a risk profile that reflected pure behavioral problems in the absence of other risk factors. Relatively high levels of externalizing problems tended to co-occur with attention-, emotional-, and social problems. Though it is difficult to draw conclusions because of the small number of children with this specific risk profile, children with these combinations of risks seem in need of additional support: they still displayed worrisome levels of psychopathology after two years of universal intervention.

Another risk group that could be qualified as non-responders encompassed at-risk children exposed to poor parenting and sociodemographic risks. These children showed modest signs of behavioral and emotional problems. The ineffectiveness of the GBG is perhaps not surprising given that parenting skills or sociodemographic risks were not targeted by the GBG. These children and families seemed in need of interventions that include parents.

In conclusion, this study takes an important step forward in establishing the effectiveness of the GBG by elucidating to whom intervention effects can be generalized (Flay et al., 2005). The GBG was most effective among children with internalizing problems and among victimized children but also prevented normative growth in behavioral and emotional symptoms among low-risk children. Importantly, the effectiveness of the GBG appears questionable when the severity and breadth of early sociobehavioral problems is extensive. Also, at-risk children from dysfunctional families could be qualified as intervention non-responders. Of note, there were no indications of unwanted or iatrogenetic effects (Flay et al., 2005).

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

The present findings have implications for practice. Knowledge of baseline characteristics at school entry that predict non-response to universal school-based prevention programs is essential for schools to develop a continuum of interventions from universal to selective to indicated prevention strategies to meet the needs of all their students (Mrazek & Haggerty, 1994; Offord, Kraemer, Kazdin, Jensen, & Harrington, 1998). Early identification may be in particular crucial for children with severe sociobehavioral risk who need more extensive intervention (cf. Eron et al., 2002). This study indicates that for the accurate identification of children and families in need for additional support (e.g., parent support programs, social skills training), early screening across multiple ecological levels including the peer, family, and demographic context is critical. This study also suggests that this could be accomplished with relatively brief parent- and teacher reports, though more research is needed to replicate findings in different samples with different universal classroom programs to validate the current approach. Furthermore, the results can guide the development and refinement of universal prevention programs for students with particular constellations of risks. In this particular case, for three of the five identified risk groups, the implemented group-based contingency program did not seem to adequately address the causal mechanisms underlying the adjustment problems, warranting further reflection on the components and focus of the GBG. A behavioral-environmental contingency program may not be effective in changing cognitive- affective processes linked with problem behavior.

This study presented a novel way to identify a-priori unknown subgroups of children that differ in intervention responsiveness. The results highlight the potential of a person-centered approach, allowing for the simultaneous assessment of both the severity and breadth of dysfunction across the child, peer, family and societal domain (cf. Lanza & Rhoades, 2011). The limitations of the study, however, should also be considered. First, the statistical approach

presents a data-driven technique. Therefore, replication in other school-based samples is warranted. Second, the classification of subgroups of children compromised the power to detect significant within and between subgroup differences due to small and unequal class sizes. Third, the duration and stability of baseline characteristics was not taken into account. Fourth, although reports on the immediate intervention outcomes were obtained from three different teachers across the study years (teachers typically change per year) these were also the ones who implemented GBG (except for the kindergarten teacher). However, there is some evidence for long-term effects of the GBG on outcomes not reported by the implementing teachers (e.g., Ialongo et al., 2001; Kellam et al., 2008). Lastly, the intervention was delivered in early grade school. Annual follow-up assessments are needed to examine whether the subgroup differences are temporary, stable or enlarge over the course of grade school. In addition, the study does not address the issue of developmental timing and duration (shorter or longer than 2 years) of universal intervention.

In summary, this study was among the first to examine the question which children are impacted by an universal school-based preventive intervention using a person-centered approach. Overall, the Good Behavior Game was particularly effective among children with internalizing problems but also among victimized children and low-risk children. The GBG seemed not effective for children exposed to risks at multiple levels, including children with sociobehavioral risks and children with family-demographic risk profiles.

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

References

Abidin, R. R. (1983). *Parenting stress: Index manual*. Charlottesville: Pediatric Psychology Press.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Ashford, J., Smit, F., Van Lier, P. A. C., Cuijpers, P., & Koot, H. M. (2008). Early risk indicators of internalizing problems in late childhood: A 9-year longitudinal study. *Journal of Child Psychology and Psychiatry, 49*, 774-780.

Barrish, H. H., Saunders, M., & Wolf, M. M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavioral Analysis, 2*, 119-124.

Beauchaine, T. P., Webster-Stratton, C., & Reid, J. (2005). Mediators, moderators, and predictors of 1-year outcomes among children treated for early-onset conduct problems: A latent growth curve analysis. *Journal of Consulting & Clinical Psychology, 73*, 371-388.

Bierman, K. (2002). Predictor variables associated with positive fast track outcomes at the end of third grade. *Journal of Abnormal Child Psychology, 30*, 37-52.

Bierman, K. L., Coie, J. D., Dodge, K. A., Greenberg, M. T., Lochman, J. E., McMahon, R. J., et al. (2010). The effects of a multiyear universal social–emotional learning program: The role of student and school characteristics. *Journal of Consulting and Clinical Psychology, 78*, 156-168

Bloom, H. S., & Michalopoulos, C. (2011). When is the story in the subgroups? Strategies for interpreting and reporting intervention effects for subgroups. *Prevention Science*. doi:10.1007/s11121-010-0198-x

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513-531.

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behavior and social status in the school: a cross-age comparison. *Child Development 59,* 815-829.

De Brock, A. J. L. L., Vermulst, A. A., Gerris, J. R. M., & Abidin, R. R. (1992). *Nijmeegse Ouderlijke Stress Index (Nijmegen Parental Stress Index).* Lisse: Swets & Zeitlinger.

Dodge, K. A., & Pettit, G. S. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65,* 649-665.

Dolan, L. J., Jaylan, T., Werthamer, L., & Kellam, S. G. (1989). *The Good Behavior Game manual*. Baltimomore, MD: The Johns Hopkins Prevention Research Center.

Erasmus Medical Center. (2000). *Problem Behavior at School Interview*. Rotterdam: Erasmus University.

Eron, L., Huesmann, R., Spindler, A., Guerra, N., Henry, D., & Tolan, P. (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high-risk children. *Journal of Consulting and Clinical Psychology, 70,* 179-194.

Farrell, A. D., Henry, D. B., & Bettencourt, A. (2011). Methodological challenges examining subgroup differences: Examples from universal school-based youth violence prevention trials. *Prevention Science*. doi: 10.1007/s11121-011-0200-2

Farrell, A. D., Henry, D. B., Mays, S. A., & Schoeny, M. E. (2011). Parents as moderators of the impact of school norms and peer influences on aggression in middle school students. *Child Development, 82,* 146-161.

Flay, B., Biglan, A., Boruch, R., Castro, F., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151-175.

Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory: An integrative approach*. Newbury Park, CA: Sage.

Gazelle, H., & Ladd, G. W. (2003). Anxious solitude and peer exclusion: A diathesis-stress model of internalizing trajectories in childhood. *Child Development, 74*, 257-278.

Greca, A. M. L., Silverman, W. K., & Lochman, J. E. (2009). Moving beyond efficacy and effectiveness in child and adolescent intervention research. *Journal of Consulting and Clinical Psychology*, *77*, 373-382.

Hawes, D. J., & Dadds, M. R. (2006). Assessing parenting practices through parent-report and direct observation during parent-training. *Journal of Child & Family Studies, 15*, 554-567.

Hinshaw, S. P. (2007). Moderators and mediators of treatment outcome for youth with ADHD: Understanding for whom and how interventions work. *Journal of Pediatric Psychology, 32*, 664-675.

Ialongo, N., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional & Behavioral Disorders, 9*, 146-160.

Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27*, 599-641.

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass 2/1 2*, 302-314.

Kahn, R. S., Brandt, D., & Whitaker, R. C. (2004). Combined effect of mothers' and fathers' mental health symptoms on children's behavioral and emotional well-being. *Archives of Pediatric and Adolescent Medicine, 158*, 721-729.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., et al. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*, 5-28.

Kessler, R. C., Andrwes, G., Colpe, L. J., Hiripi, E. M., Roczek, D. K., Normand, S.-L. T., et al. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*, 959-976.

Ladd, G. W. (2006). Peer rejection, aggressive or withdrawn behavior, and psychological maladjustment from ages 5 to 12: An examination of four predictive models. *Child Development, 77*, 822-846

Lanza, S., & Rhoades, B. (2011). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science.* doi:10.1007/s11121-011-0201-1

Lochman, J. E. (2004). Contextual factors in the risk and prevention research. *Merrill-Palmer Quarterly, 50*, 311-325.

Lochman, J. E. (2006). Translation of research into interventions. *International Journal of Behavioral Development, 30*, 31-38.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology, 48*, 371-410.

Mrazek P, & Haggerty R. (1994). *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention.* Washington, DC: National Academy Press.

McKee, L., Forehand, R., Rakow, A., Reeslund, K., Roland, E., Hardcastle, E., et al. (2008). Parenting specificity: An examination of the relation between three parenting behaviors and child problem behaviors in the context of a history of caregiver depression *Behavior Modification, 32,* 638-658.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100,* 674-701.

Muthén, B., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882-891.

Muthén, L. K., & Muthén, B. O. (1998-2007). Mplus (Vers. 5). Los Angeles: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Offord, D. R., Kraemer, H. C., Kazdin, A. E., Jensen, P. S., & Harrington, R. (1998). Lowering the burden of suffering from child psychiatric disorder: Trade-offs among clinical, targeted, and universal interventions. *Journal of the American Academy of Child & Adolescent Psychiatry, 37,* 686-694.

Parker, J., Rubin, K. H., Erath, S., Wojslawowicz, J. C., & Buskirk, A. A. (2006). Peer relationships and developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental Psychopathology: Risk, Disorder, and Adaptation* (2nd ed., Vol. 2, pp. 419-493). New York: Wiley.

Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist, 44*, 329-329-335.

Petras, H., Kellam, S. G., Brown, C. H., Muthén, B. O., Ialongo, N. S., & Poduska, J. M. (2008). Developmental epidemiological courses leading to antisocial personality disorder and violent and criminal behavior: Effects by young adulthood of a universal preventive intervention in first- and second-grade classrooms. *Drug and Alcohol Dependence, 95*, S45-S59.

Shaw, D. S., Bell, R. Q., & Gilliom, M. (2000). A truly early starter model of antisocial behavior revisited. *Clinical Child and Family Psychology Review, 3*, 155-172

Stevens, G. W. J. M., & Vollebergh, W. A. M. (2008). Mental health in migrant children. *Journal of Child Psychology and Psychiatry, 49*, 276-294.

The Multisite Violence Prevention Project (2008). The multisite violence prevention project: Impact of a universal school-based violence prevention program on social-cognitive outcomes. *Prevention Science, 9*, 231-244.

van Lier, P. A. C., Muthén, B. O., van der Sar, R. M., & Crijnen, A. (2004). Preventing disruptive behavior in elementary school children: Impact of a universal classroom-based intervention. *Journal of Consulting & Clinical Psychology, 72*, 467-478.

van Lier, P. A. C., Vuijk, P., & Crijnen, A. (2005). Understanding mechanisms of change in the development of antisocial behavior: The impact of a universal intervention. *Journal of Abnormal Child Psychology, 33*, 521-535.

Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1997). *Handleiding voor de Teacher's Report Form.* Rotterdam: Erasmus Universiteit Rotterdam.

Witvliet, M., Lier, P. A. C. v., Cuijpers, P., & Koot, H. M. (2009). Testing links between childhood positive peer relations and externalizing outcomes through a randomized controlled intervention study. *Journal of Consulting & Clinical Psychology, 77*, 905-915.

Tables

Table 1. *Correlations between Continuous Risk and Outcome Variables at Pre- (Wave 1) and Postintervention (Wave 4)*

	M (SD)	1	2	3	4	5	6	7	8	9	10	11
Risk variables												
1. Internalizing problems	0.15 (.19)											
2. Externalizing problems	0.16 (.25)	.32**										
3. Attention problems	0.25 (.30)	.37**	.78**									
4. Social problems	0.14 (.20)	.54**	.66**	.65**								
5. Victimization	0.59 (.57)	.01	.32**	.31**	.25**							
6. Parenting stress	0.90 (.73)	.07	.10*	.13**	.06	.13**						
7. Depression	1.48 (1.85)	.01	.17**	.22**	.09	.02	.31**					
8. Involvement	2.89 (.60)	-.04	-.05	-.08	-.04	-.05	-.28**	-.11*				
Outcome variables												
9. Externalizing-Wave 1	0.78 (.67)	.11**	.73**	.62**	.48**	.29**	.16**	.12*	-.08			
10.Externalizing-Wave 4	0.74 (.70)	.02	.51**	.50**	.31**	.31**	.14**	.22**	-.15**	.52**		
11.Internalizing-Wave 1	0.79 (.57)	.53**	.14**	.26**	.35**	.01	.04	-.03	-.00	.33**	.05	
12.Internalizing-Wave 4	1.01 (.73)	.19**	.08	.16**	.14**	.14**	.06	.09	-.03	.05	.39**	.20**

Note 1: $*p \leq .05$, $**p \leq .01$ (two-tailed)

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

Table 2. *Effect Size of the Difference in Observed Mean Scores between Control and GBG after Two Years of Intervention for Each Risk Profile*

	Low risk (*n*=396)	Victimization risk (*n*=107)	Internalizing risk (*n*=60)	Family and demographic risk (*n*=106)	Moderately-high sociobehavioral risk (*n*=51)	Severe sociobehavioral risk (*n*=21)
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Externalizing						
Control	0.63 (.52)	0.81 (.70)	1.38 (.69)	1.38 (.68)	1.57 (.57)	2.66 (.71)
GBG	0.43 (.44)	0.33 (.39)	0.39 (.45)	1.26 (.61)	1.33 (.71)	2.13 (.70)
Effect size	0.42	0.85	1.70	0.19	0.37	0.75
Internalizing						
Control	1.15 (.69)	1.25 (.60)	2.02 (.51)	1.25 (.76)	0.98 (.58)	1.54 (.07)
GBG	0.81 (.67)	0.59 (.61)	1.06 (.69)	1.11 (.71)	0.96 (.75)	1.21 (.72)
Effect size	0.50	1.09	1.58	0.19	0.03	0.65

Note: Effect size = Cohen's *d*

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

Table 3. *Growth Parameters and Estimates of Intervention Effects for Each Risk Profile*

	Low risk (n=396)	Victimization risk (n=107)	Internalizing risk (n=60)	Family and demographic risk (n=106)	Moderately-high sociobehavioral risk (n=51)	Severe sociobehavioral risk (n=21)
	Est. (SE)	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*	*Est. (SE)*
Externalizing						
Intercept	0.35 (.03)***	1.00 (.07)***	0.70 (.08)***	1.24 (.09)***	1.78 (.09)***	2.48 (.09)***
Slope	0.16 (.04)*	-0.08 (.21)	0.28 (.10)**	0.06 (.07)	-0.20 (.11)	-0.05 (.19)
Slope on GBG	-0.11 (.05)* $_{a}$	-0.28 (.21) $_{a\,b}$	-0.41 (.10)*** $_{b}$	-0.09 (.12) $_{a}$	-0.06 (.14) $_{a}$	-0.09 (.23) $_{a\,b}$
Internalizing						
Intercept	0.54 (.04)***	1.13 (.10)***	1.34 (.09)***	0.76 (.07)***	0.91 (.09)***	1.30 (.12)***
Slope	0.33 (.05)***	0.02 (.16)	0.35 (.07)***	0.25 (.10)*	0.09 (.14)	0.13 (.11)
Slope on GBG	-0.18 (.07)** $_{a\,b}$	-0.31 (.14)* $_{b}$	-0.50 (.09)*** $_{c}$	-0.09 (.11) $_{a}$	-0.10 (.15) $_{a\,b}$	-0.02 (.20) $_{a\,b}$

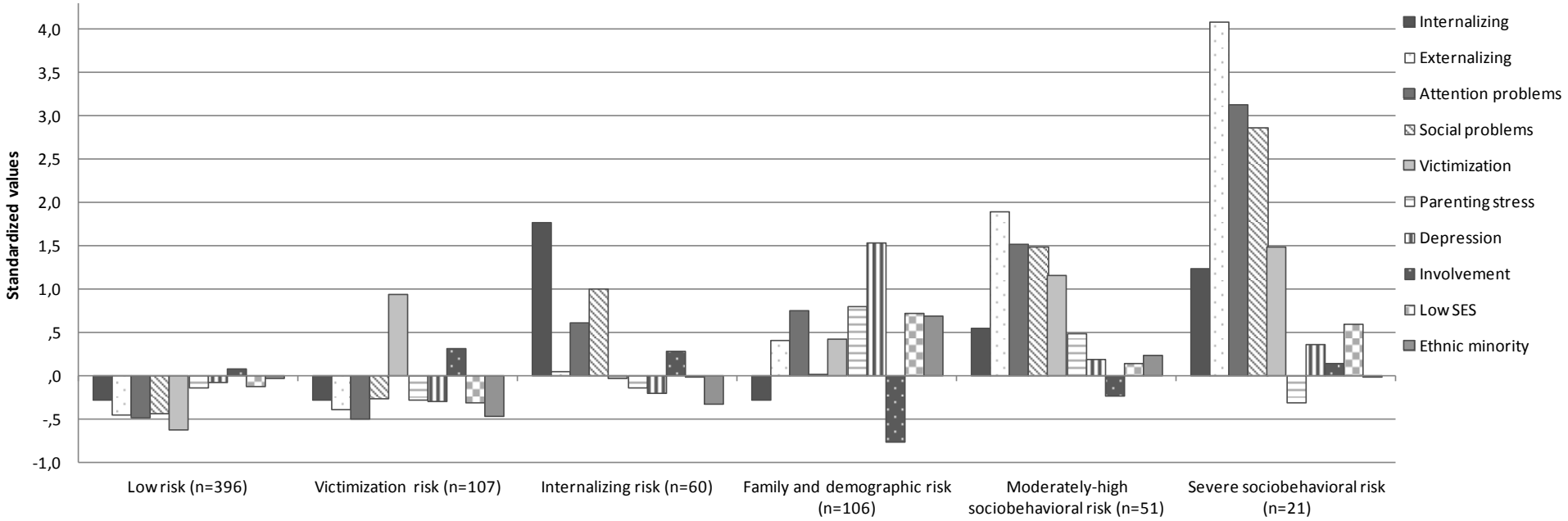
Note 1: $*p \leq .05$, $**p \leq .01$, $***p \leq .001$ (two-tailed)

Note 2: Intervention effects (Slope on GBG) that share the same subscript (a, b, c) are similar (two-tailed, $p \leq .05$). Marginally significant ($p \leq .08$) differences in intervention effects on Internalizing behavior not displayed in the table were found between the *Low risk group* and the *Victimization risk group*, and between the *Victimization risk group* and the *Moderately-high sociobehavioral risk group*.

This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7

Figures

Figure 1. *Baseline Risk Profiles*



This is an Author's Accepted Manuscript of an article published in Prevention Science, 2013, 14, 479-488. © Springer, the final publication is available online at http://dx.doi.org/10.1007/s11121-012-0329-7